Low Level Design (LLD)

Concrete Compressive Strength Prediction (Machine Learning)

By

Kavitha Narsapur

Abstract

The quality of concrete is determined by its compressive strength, which is measured using a conventional crushing test on a concrete cylinder. The strength of the concrete is also a vital aspect in achieving the requisite longevity. It will take 28 days to test strength, which is a long period. Thus this project aims to predict concrete strength using machine learning models. Different models like Linear regression, Lasso regression, Decision tree regression, Random forest regression, Ada boost regression, Gradient boost regression, XGBoost regression and Voting regression were built. Out of all the regressors, XGBoost regression resulted in highest R2_score of 92.4635% and low RMSE of 4.787508. Hence XGBoost regression can be used for predicting concrete compressive strength.

Introduction

Why this Low-Level Design Document?

This Low-Level Design (LLD) Document summarizes the data used for predicting concrete compressive strength and also provides an overview of steps used in this project as Data preprocessing, Exploratory data analysis, Model building and Model deployment.

Scope

The LLD documentation presents the structure of the system, such as the application architecture (layers), application flow (Navigation), and technology architecture. The LLD uses non-technical to mildly-technical terms which should be understandable to the administrators of the system. This software system will be a Web application. This system will be designed to predict concrete compressive strength.

Dataset overview

Number of instances: 1030

Number of Attributes: 9

Attribute breakdown 8 quantitative input variables, and 1 quantitative output variable

Missing Attribute Values: None

Given are the variable name, variable type, the measurement unit and a brief description. The concrete compressive strength is the regression problem. The order of this listing corresponds to the order of numerals along the rows of the database.

Name -- Data Type -- Measurement -- Description

Cement (component 1) -- quantitative -- kg in a m3 mixture -- Input Variable

Blast Furnace Slag (component 2) -- quantitative -- kg in a m3 mixture -- Input Variable

Fly Ash (component 3) -- quantitative -- kg in a m3 mixture -- Input Variable

Water (component 4) -- quantitative -- kg in a m3 mixture -- Input Variable

Superplasticizer (component 5) -- quantitative -- kg in a m3 mixture -- Input Variable

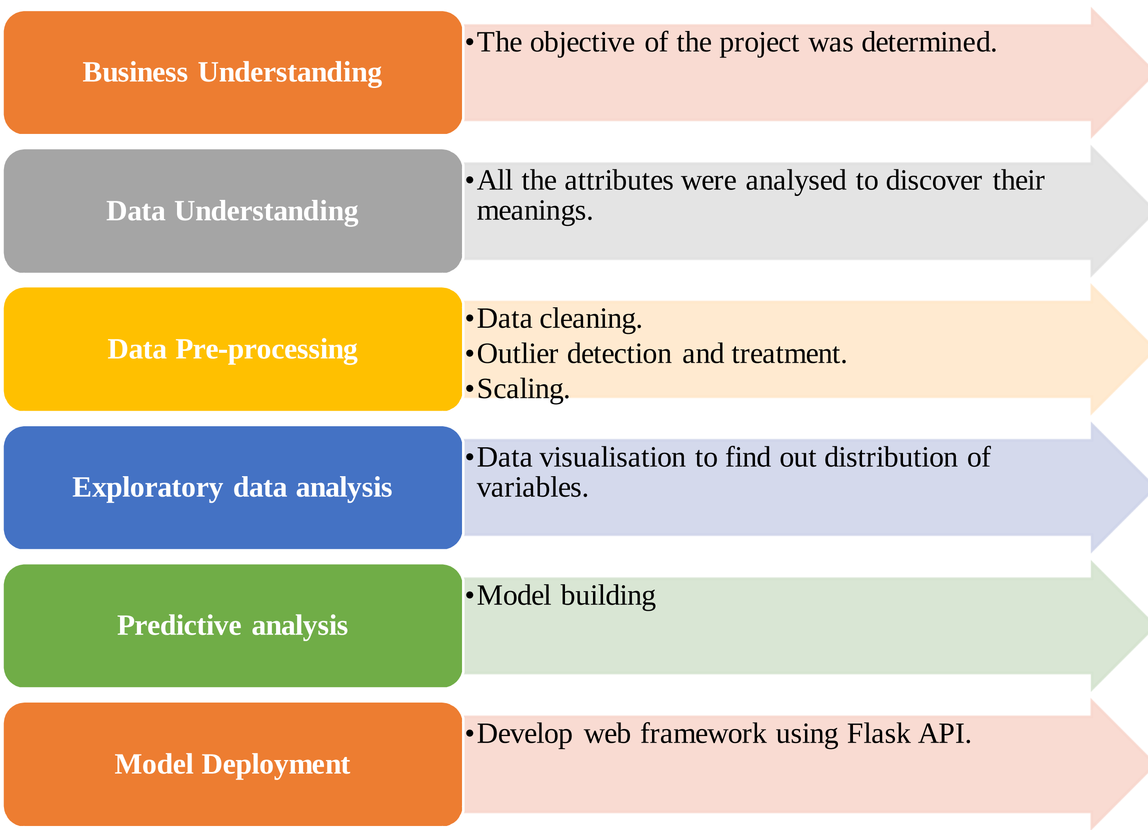
Coarse Aggregate (component 6) -- quantitative -- kg in a m3 mixture -- Input Variable

Fine Aggregate (component 7) -- quantitative -- kg in a m3 mixture -- Input Variable

Age -- quantitative -- Day (1~365) -- Input Variable

Concrete compressive strength -- quantitative -- MPa -- Output Variable

Architecture



Architecture Description

1. Business objective:

The aim of this project is to develop a solution using Data science and machine learning to predict the compressive strength of a concrete with respect to the its age and the quantity of ingredients used.

2. Data Understanding:

An attempt was made to understand the meanings of all the variables present in the data. The data type of each variable was also determined.

3. Data pre-processing:

Data cleaning: Data was checked for the presence of missing values.

Outlier detection and treatment: Boxplots were used to detect outliers. Outliers were then treated by the method of capping and flooring using Interquartile range (IQR).

Scaling: All the attributes were brought to same scale using standard scaler.

4. Exploratory data analysis:

By means of graphical analysis, distribution of every variable was determined.

5. Predictive analysis:

Train and Test data creation: Train and Test data were created by splitting the data and 70% of the data was used as train data and 30% was used as test data.

Model building: Different models like Linear regression, Lasso regression, Decision tree regression, Random forest regression, Ada boost regression, Gradient boost regression, XGBoost regression and Voting regression were built.

Model evaluation: The built models were tested on the test data and evaluation metrics used were R2_Score and RMSE.

6. Model Deployment:

Best performing model was saved in Pickle format and the model was tested using Flask API on local system.

Conclusion

The web framework developed in this project can be used by the users to predict the concrete compressive strength.